

Rapport Projet 4

Sylvain VASSEUR



16 avril 2025

Projet personnel à but formateur et exploratoire

Table des matières

[I. Résumé exécutif 2](#_Toc195875952)

[II. Introduction 2](#_Toc195875953)

[III. Méthodologie 3](#_Toc195875954)

[IV. Résultats 4](#_Toc195875955)

[V. Analyse approfondie 6](#_Toc195875956)

[VI. Recommandations 6](#_Toc195875957)

[VII. Conclusion 6](#_Toc195875958)

# Résumé exécutif

* **Contenu**: Contexte, objectif, méthodes, résultats clés, recommandations principales
* **Bonnes pratiques**: Limiter à 1 page, être percutant, centrer sur la valeur business

# Introduction

* **Contenu**: Contexte business, objectifs précis, questions de recherche

Nous souhaitons analyser les écarts de performance entre ses départements.

* **Bonnes pratiques**: Formuler clairement les questions auxquelles l'analyse répond

Contexte  
Dans un souci d’amélioration continue, une entreprise a exprimé le besoin d’évaluer les écarts de performance entre ses différents départements. Cette initiative vise à mieux comprendre les facteurs internes qui influencent la productivité et la satisfaction des collaborateurs.

Objectifs  
L’objectif principal de cette analyse est d’identifier les différences de performance entre départements, à partir de variables à la fois qualitatives (comme le type de département ou le niveau de satisfaction) et quantitatives (comme les scores de performance ou les heures travaillées). Il s’agit également de détecter d’éventuelles corrélations ou tendances permettant d’orienter des décisions de gestion ou d’optimisation des ressources.

Questions de recherche

* Existe-t-il des écarts de performance significatifs entre les départements ?
* Quels facteurs (satisfaction des employés, heures travaillées) expliquent les écarts de performance observés entre les départements ?
* Quelles relations spécifiques peut-on identifier entre les variables qualitatives (départements, satisfaction) et quantitatives (scores de performance, heures travaillées) ?
* Peut-on identifier des profils de départements plus performants ou moins performants que la moyenne ?

# Méthodologie

* **Contenu**: Sources de données, préparation, méthodes d'analyse, limites
* **Bonnes pratiques**:
  + Être concis mais précis
  + Expliquer les choix méthodologiques
  + Mentionner les tests statistiques utilisés
  + Décrire les transformations de données

##### Analyse des départements :

**Méthodes :** Statistiques descriptives comparatives, tests d'hypothèses sur les différences (ANOVA, tests t), analyse de distribution des performances

**Objectif :** Identifier et quantifier les écarts de performance entre les départements, en répondant à la question : *"Existe-t-il des écarts significatifs de performance entre les départements ?"*

##### Analyse des facteurs explicatifs :

**Méthodes :** Analyse de corrélation bivariée et multivariée, modélisation des relations entre variables, analyse de régression

**Objectif :** Déterminer l'influence de la satisfaction des employés et des heures travaillées sur la performance, pour répondre à : *"Quels facteurs expliquent les écarts de performance observés entre les départements ?"*

##### Identification des profils de départements :

**Méthodes :** Techniques de segmentation, analyse comparative des groupes de performance, caractérisation multidimensionnelle des profils

**Objectif :** Caractériser les départements surperformants et sous-performants et identifier leurs attributs distinctifs, répondant à : *"Peut-on identifier des profils de départements plus performants ou moins performants que la moyenne ?"*

##### **Synthèse et recommandations :**

**Méthodes :** Analyse croisée des résultats, hiérarchisation des facteurs d'influence, modélisation d'impact des interventions potentielles

**Objectif :** Formuler des recommandations concrètes et hiérarchisées pour réduire les écarts de performance entre départements.

# Résultats

* **Contenu**: Statistiques descriptives, observations factuelles, tendances principales
* **Bonnes pratiques**:
  + Présenter les faits sans interprétation approfondie
  + Utiliser des graphiques pour illustrer les données brutes
  + Organiser par thème/question de recherche
  + Mettre en évidence les observations surprenantes

1. Analyse des départements
   1. Analyse des statistiques descriptives

Cette section présente les résultats de l'analyse quantitative des écarts de performance entre départements. Les données ont été analysées à l'aide de statistiques descriptives et de tests d'inférence statistique conformément à la méthodologie décrite précédemment.

Dans un premier temps, les indicateurs de tendance centrale et de dispersion sont examinés pour chaque département. Le Tableau 1 synthétise ces statistiques fondamentales, permettant une première comparaison factuelle des performances observées.

[TABLEAU 1] Tableau stats descriptives

Une remarquable **homogénéité** ressort des données : les moyennes varient dans un intervalle restreint, de 2,98 à 3,02, soit une amplitude de 0,04 points. De même, les écarts-types présentent une variabilité minimale (1,41 à 1,42). Les valeurs minimales et maximales sont identiques pour toutes les catégories, et le nombre d'observations par catégorie s'élève à environ 11 000 unités.

Ces résultats indiquent une **distribution** des performances **très uniforme** entre les neuf catégories, sans écart notable entre elles.

* 1. Tests d’hypothèses sur les différences

Afin d’identifier d’éventuelles disparités entre les départements, un test d’ANOVA est réalisé.  
Ce test permet de détecter si, pour une variable donnée, **au moins deux départements** présentent une **différence significative de moyenne**.

Les variables sélectionnées pour cette analyse sont les suivantes :

* Work\_Hours\_Per\_Week : volume horaire de travail
* Overtime\_Hours : charge ou pression supplémentaire
* Projects\_Handled : nombre de missions gérées
* Performance\_Score : niveau de performance évalué
* Remote\_Work\_Frequency *:* fréquence de télétravail
* Team\_Size *:* effectif moyen par équipe
* Monthly\_Salary *:* niveau de rémunération mensuelle

Ces variables ont été choisies car elles permettent de **caractériser de façon concrète l'organisation du travail dans chaque département**. Un résultat significatif à l’ANOVA indiquerait qu’au moins deux départements présentent des **conditions ou résultats sensiblement différents** sur l’un de ces aspects — ce qui pourrait orienter des décisions de pilotage ou d’équilibrage interne.

[TABLEAU 2] Tableau du test ANOVA

Ces résultats suggèrent qu'il n'y a pas de disparité significative entre les départements en termes de **charge de travail**, **pression supplémentaire**, **nombre de projets gérés**, **performance**, **télétravail**, **taille des équipes**, et **salaires**.

* 1. Analyse de distribution des performances

Bien que certaines variables présentent des p-values proches du seuil de significativité (notamment Projects\_Handled avec p = 0.086), ces différences ne sont pas assez marquées pour être considérées comme statistiquement significatives au seuil classique de 5%. Par conséquent, ces résultats ne justifient pas une distinction notable entre les départements pour les aspects analysés.

**Intéressons-nous à la répartition des scores de performance au sein des départements.** L’analyse statistique suggère que les distributions des scores sont, à quelques nuances près, similaires entre les départements.

[TABLEAU 3] Histogrammes % perf by departemetn

Cette observation est confirmée par les données : la répartition des scores de performance ne montre pas de disparités significatives entre les départements. La proportion d’employés pour chaque niveau de score (de 1 à 5) reste globalement stable d’un secteur à l’autre.

* 1. Synthèse des observations

L’analyse des scores de performance selon les départements révèle une forte similarité entre les groupes. Les statistiques descriptives montrent des moyennes très proches et des écarts types quasi identiques, le tout avec un volume de données similaire entre départements. Les résultats du test ANOVA confirment cette stabilité, aucune des variables testées ne présentant de différence significative entre les groupes. Par ailleurs, la répartition des scores de performance est équilibrée dans l’ensemble des départements, chaque modalité représentant environ 20 % des effectifs. Ces constats ne mettent en évidence **aucune variation notable liée au département à ce stade de l’analyse**. Pour la suite, nous pourrons nous concentrer sur d’éventuels leviers internes aux départements ou sur des variables transverses susceptibles d’apporter un éclairage complémentaire.

1. **Analyse des facteurs explicatifs**
   1. Analyse des corrélations du score de Performance

Afin de mieux comprendre quels facteurs peuvent influencer le *Performance\_Score*, nous avons réalisé une analyse de corrélation. L’objectif est d’identifier si certaines variables sont significativement liées à la performance des employés.

[TABLEAU 3] Corrélation de Perforamnce Score

Les résultats sont globalement peu concluants : à l’exception d’une seule variable, toutes les corrélations obtenues sont proches de zéro. Cela signifie qu’aucun lien clair n’émerge entre le score de performance et la majorité des variables analysées. Ces résultats freinent notre capacité à dégager des pistes concrètes d’amélioration.

Seule exception notable : le Monthly\_Salary, qui présente une corrélation de **0.51** avec le Performance\_Score. Cette valeur suggère une relation modérée — autrement dit, plus un employé est rémunéré, plus son score de performance tend à être élevé. Ce lien reste cependant à nuancer : bien qu’une corrélation de 0.51 soit intéressante, elle ne permet pas, à elle seule, d’expliquer pleinement les différences de performance observées.

Afin d’avoir une vision globale de l’impact des variables quantitatives, nous réalisons une régression linéaire multiple, qui permet d’évaluer l’effet combiné de plusieurs variables sur une variable cible – ici, l’**indice de performance**.

**[TABLEAU 3] — Régression multivariée du Performance\_Score**

Nous nous intéressons uniquement aux variables dont la **valeur p** est inférieure à 0.05, ce qui signifie que leur influence est statistiquement significative. Dans notre cas, seules deux variables remplissent ce critère : Monthly\_Salary et Team\_Size.

Le coefficient associé au salaire mensuel peut paraître faible (0.0005), mais il s’applique à une variable aux valeurs élevées. Ainsi, une variation réaliste du salaire entraîne une variation notable du score de performance. Ce type de lecture rappelle que la signification d’un coefficient dépend autant de sa valeur que de l’échelle de la variable concernée.

De son côté, **Team\_Size** affiche un coefficient négatif (-0.0017), ce qui suggère qu’à mesure que la taille de l’équipe augmente, le score de performance individuel tend à diminuer légèrement.

Enfin, le **coefficient de détermination** R² = 0.26 indique que l’ensemble de nos variables explicatives parvient à expliquer **26% de la variance observée** du score de performance. Cela reste une proportion honorable dans un contexte RH, où de nombreux facteurs externes ou qualitatifs (non mesurés ici) peuvent influencer la performance.

[TABLEAU 3] Corrélation de Perforamnce Score selon les departements

* 1. Sdf
  2. Résultats d’analyse

Regarder la corrélation du salaire mensuel car c’est la variable qui influe le plus sur l’indice de performance

Plan

### **B. Analyse des facteurs explicatifs des écarts de score entre départements**

**B.1. Analyse globale des corrélations (tous départements confondus)**

Objectif : **identifier les variables liées au score**, sans distinction de département.  
Tu testes toutes les variables non encore analysées (âge, ancienneté, projets, etc.) par rapport au score.  
Tu repères les liens forts ou faibles, sans interpréter, juste pour **dégager des variables d’intérêt** pour la suite.

* Corrélation de toutes les variables avec le score
* Visualisations possibles (heatmap, scatter, etc.)
* Repérage de variables potentiellement explicatives

**B.2. Analyse différenciée selon les départements**

Objectif : **voir si les facteurs explicatifs varient selon les départements**, donc s’il y a des comportements internes différents.  
Tu prends les **variables identifiées en B.1**, et tu refais les corrélations **par département**.  
=> Tu compares les corrélations entre départements pour voir s’il y a des spécificités locales.

* Corrélations score / variable 1 dans chaque département
* Idem pour variable 2, etc.
* Visualisation type : facet grid, heatmap par département, etc.
* Tu ne tires pas de conclusion, tu constates les différences ou similarités

**B.3. Modélisation par département (régressions locales)**

Objectif : **approfondir** en construisant un modèle de score **par département**, pour voir **quelles variables ont un effet propre dans chaque contexte**.  
Tu fais une régression multiple par département avec les variables les plus pertinentes (tu peux les choisir selon B.2).  
Tu obtiens donc 9 modèles distincts, chacun avec ses propres coefficients.  
Tu ne cherches pas à interpréter les coefficients (sauf si demandé plus tard), tu observes s’ils varient ou non.

* Régression du score par département
* Présentation synthétique des coefficients (tableau, graphique type “coefficient plot”)
* Comparaison globale des poids des variables entre départements

# Analyse approfondie

* **Contenu**: Interprétation des résultats, relations causales, facteurs explicatifs
* **Bonnes pratiques**:
  + Établir des liens entre différentes observations
  + Approfondir les éléments les plus significatifs
  + Tester et valider les hypothèses
  + Éviter de répéter les résultats, les référencer

# Recommandations

* **Contenu**: Actions concrètes, priorisation, mesures de suivi
* **Bonnes pratiques**:
  + Lier directement aux analyses
  + Quantifier l'impact attendu quand possible
  + Prioriser par impact/effort

# Conclusion

* **Contenu**: Synthèse, limites, perspectives futures
* **Bonnes pratiques**: Suggérer les prochaines pistes d'analyse